LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-BOOK-402916

Systems biology of Microbial Communities

A. Navid, C.-M. Ghim, A. Fenley, S. Yoon, S. Lee, E. Almaas

April 14, 2008

Systems biology

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Systems biology of microbial communities

Ali Navid, Cheol-Min Ghim, Andrew Fenley, Sooyeon Yoon, Sungmin Lee, Eivind Almaas*

Biosciences & Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Abstract

Microbes exist naturally in a wide range of environments, spanning the extremes of high acidity and high temperature to soil and the ocean, in communities where their interactions are significant. We present a practical discussion of three different approaches for modeling microbial communities: rate equations, individual-based modeling, and population dynamics. We illustrate the approaches with detailed examples. Each approach is best fit to different levels of system representation, and they have different needs for detailed biological input. Thus, this set of approaches is able to address the operation and function of microbial communities on a wide range of organizational levels.

Key Words: Microbial community, rate equation, agent-based modeling, population dynamics, quorum sensing, biofilm.

* Corresponding Author: E. Almaas <almaas@llnl.gov>

1. Introduction

Microorganisms contribute a considerable fraction of the living biomass on Earth. While traditional studies of microbes have been based on the isolation and laboratory cultivation of pure species, relatively little is known about an estimated >99% of environmental microbes due to their difficulty of cultivation under standard laboratory conditions. In fact, the vast majority of microbes naturally only occurs and thrives when in microbial *communities*: there is frequently a synergistic partitioning of metabolic function between different microbial species (Ram, 2005). The recent development of techniques to probe microorganisms in their natural environments, such as metagenomic sequencing, has uncovered an unanticipated level of phylogenetic diversity and valuable insights into lifestyle and metabolic capabilities of microbial communities occupying a broad range of environmental niches (Breitbart, 2002; Venter, 2004; Tyson, 2004).

The function and operation of microbial communities has received significant interest with the introduction of these new technologies. There is also a growing realization that microbes contribute extensively to important environmental questions such as carbon sequestration and nitrogen cycling. It has been proposed that microbes and microbial communities may provide novel avenues for the degradation of lignocellulosic material and, thus, the generation of biofuels. Recently, new findings indicate that the activity and composition of microbial communities in e.g. the intestine is of direct relevance to human obesity (Turnbaugh 2006), and revisiting the activity of pathogens, such as *Vibrio cholera*, from the community context has revealed surprising insight with immediate consequences for generating clean drinking water (Colwell 2003). Questions related to

the function and interaction of microbial consortia has therefore taken a place of prominence in the current science literature.

In this chapter, we will address three methods that have proven useful in modeling the behavior of microbial communities. These methods have different requirements for the level of detail needed to model a multi-cellular microbial system. The first method we will discuss is based on representing a microbe by rate equations, requiring the highest level of detail. Not surprisingly, this approach has been limited in applicability due to the lack of measured kinetic parameters. However, it seems plausible that this drawback will be significantly tempered in the near future. We will then describe individual-based approaches (often called agent-based modeling) capable of simulating the interaction of multiple microbes with a relatively narrow set of variables. While the focus is still on the individual microbes, this method is capable of addressing the spatial aggregation of large populations. We will complete this chapter with a discussion of population dynamics modeling, a method for which the species is the focal point. This class of approaches has become well known through the Lotka-Volterra representation of a predator prey system (Lotka 1925; Volterra 1926).

2. Rate-equation models

2.1 Background

For good reasons, all developed genome-level models of microbial metabolism are based on the assumption that the system is at steady-state (see the chapter on Flux-Balance Analysis). Although steady-state models (SSM) have shown great utility for assessing the metabolic capabilities of an organism, they ignore a number of crucial details needed to attain greater insights into the dynamics of a cell. For example:

- After an environmental or genetic perturbation, SSM only characterize the new steady-state. SSM do not calculate how long it will take for the system to reach the new steady-state and visited intermediary states.
- SSM ignore enzymatic capacity and thus cannot identify rate-limiting steps and metabolic bottlenecks.
- SSM do not account for the concentration of intermediates and thus cannot predict deleterious buildup of toxic metabolites.

Development of genome-scale kinetic models can overcome these failings; however, currently such undertakings are impractical. In order to develop a kinetic model of cellular metabolism, we must account for the time-dependent changes in metabolite concentrations. This requires the knowledge of a large number of kinetic parameters. Unfortunately, while recently developed analytical tools have accelerated genetic and proteomic analyses immensely, measurements of enzymatic kinetic parameters are still tedious and time consuming.

Kinetic models are usually developed only for well-studied pathways, such as central carbon metabolism in *Escherchia coli* (Chassangole et al. 2002), Urea cycle in *Rattus norvegicus* (Maher et al. 2003), and glycolysis in a variety of organisms ranging from single cell organisms such as *Saccharomyces cerevisiae* (e.g. Selkov 1968, Teusink et al. 2000, Hynne et al. 2001, Zhadnov& Kasemo 2001, for a review see Klipp 2007) and *Trypanosoma brucei* (Bakker et al. 2000, Navid & Ortoleva 2004) to cells from organs such as skeletal muscle (Smolen 1995) and pancreatic β-cells (Westermark & Lansner 2003). Despite their limited metabolic scope, these models have been invaluable in enhancing our understanding of the complex collective dynamics of cellular groupings.

2.2 Theory and Methodology

Perhaps the most important question that one should consider prior to developing a kinetic model is: "How detailed should the model be?" The answer to this question is directly related to other questions that have to be answered early in the modeling process. For example:

- What kinetic parameters are available?
- Is it possible to bypass or generalize certain details of a pathway and still develop a sufficiently predictive model (see e.g. Dano et al. 2006)?
- Which reactions are reversible and which are irreversible?
- Which metabolites can be transported across the cellular membrane? Are these processes passive or active (i.e. require energy expenditure)? Are they facilitated by chaperones or are they non-facilitated?
- What is the volume and surface area of a cell, and should the model account for changes to these physical characteristics?

These questions can be answered through a thorough examination of the available literature, and searching through databases such as BRENDA (www.brenda-enzymes.info). These answers will also determine how the dynamics of metabolic reactions are formulated.

Developing kinetic models of metabolic pathways involves writing the concentration changes of each metabolite as ordinary differential equations (ODEs). For example, given the two metabolic reactions:

$$A + B \xleftrightarrow{Q} \mathrm{C} \quad \text{and} \quad C \xrightarrow{k_2} D$$

where Q is the equilibrium constant equal to the ratio of forward and reverse reaction coefficients (k_1/k_{-1}). The change in concentration of the metabolites is written as:

$$\frac{d[A]}{dt} = -k_1[A][B]$$
$$\frac{d[B]}{dt} = -k_1[A][B]$$
$$\frac{d[C]}{dt} = k_1[A][B] - (k_{-1} + k_2)[C]$$
$$\frac{d[D]}{dt} = k_2[C].$$

By thus writing and solving similar ODEs for all metabolites, we may monitor the dynamic changes that occur in a microbe.

2.3 Examples

2.3.1 Coupling of glycolytic oscillations in *Saccharomyces cerevisiae*

One of the most studied problems of cellular nonlinear dynamics has been the coupling and synchronization of metabolic oscillators, such as the Baker/Brewer's yeast, *Saccharomyces cerevisiae*. The initial reports of glycolytic oscillation with a frequency of several minutes in cell-free extracts of yeasts date back to 1964 (Chance et al. 1964). Prolonged oscillations in biochemical systems require that at least one of the reactions obey nonlinear kinetics. Thus, it is not surprising that asynchronous (Markus & Hess 1984, Wolf & Heinrich 1997) and even chaotic (Markus et al. 1985, Goldbetter 1996) dynamics have been proposed. A large number of theoretical studies have examined oscillatory behavior in glycolysis, particularly in yeasts (for a review see Patnaik 2003).

The majority of theoretical studies involve the coupling of only a few metabolic pathways (eg. Wolf & Heinrich 1997a,b, Zhadnov and Kasemo 2001a), and the interaction is mediated through a common extracellular pool of metabolites that can be

imported to and/or exported from different cells. In the case of *S. cerevisiae* suspensions, acetaldehyde (Acld) has been identified as the primary coupling metabolite (Richard 1996).

2.3.1.1 Glycolysis oscillations and synchronization

The model of glycolysis in yeast was designed with the criterion that it should describe the observed experimental observations (Richard et al. 1996a,b). A schematic of the modeled system is presented in Figure 1. The characteristics of the model are:

- Several chemical steps are lumped together, such as reactions catalyzed by hexokinase and phosphofructokinase (PFK) (v_1) and multi-step conversions of dihydroxyacetone phosphate to glycerol (Gly) (v_3) and 3-phosphoglycerate (3PG) to pyruvate (Pyr) (v_7).
- Simulation is for anaerobic conditions with ethanol (Etoh) as the major product.
- Concentrations of Gly and Etoh are considered constant (reservoir). Import of glucose and export of Acld are the only modeled extracellular fluxes: Import of glucose (I) is assumed constant, and transport of Acld (X) is modeled as passive diffusion dependent on the concentration gradient of Acld across the membrane:

$$X = \frac{AJ}{V}([Acld(c)] - [Acld(m)])$$

  where A and V are the surface area and volume of the cell respectively. J is the coefficient of permeability of the cellular membrane for Acld. c and m denote cytosolic and medium concentrations.
- Consumption of ATP by the cell is accounted for by an ATPase reaction. Intracellular pools of adenine nucleotides (ATP and ADP) and nicotinamide

adenine dinucleotides (NAD^+ and NADH) are conserved:

$$[ATP] + [ADP] = A_{Total} \text{ and } [NAD^+]+[NADH]=N_{Total}.$$

- All reaction considered irreversible, except for glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (v_4) and phosphoglycerate kinase (PGK) (v_5).
- Reactions catalyzed by GAPDH and PGK are near equilibrium (Q_{GAPDH}=0.0056, Q_{PGK}=3225 (Bergmeyer 1974)), justifying a quasi-steady-state approximation for 1,3-bisphosphoglycerate (13BPG), thus $\frac{d[13BPG]}{dt} = 0$. And since,

$$\frac{d[13BPG]}{dt} = v_4 - v_5,$$

$$v_4 = k_4[TRP][NAD^+] - k_{-4}[13BPG][NADH],$$

$$v_5 = k_5[13BPG][ADP] - k_{-5}[3PG][ATP]$$

we can write the equation for concentration of 13BPG as

$$[13BPG] = \frac{k_4[TRP][NAD^+] + k_{-5}[3PG][ATP]}{k_{-4}[NADH] + k_5[ADP]}.$$

Thus, the combined reaction equation for v_4 and v_5 becomes:

$$v_{4\&5} = \frac{k_4k_5[TRP][NAD^+](A_{Total} - [ATP]) - k_{-4}k_{-5}[3PG][ATP][NADH]}{k_{-4}[NADH] + k_5[ADP]}.$$

- Simple rate laws are used for all enzymatic reactions (see Table 1).
- The only regulatory behavior that is accounted for is the inhibitory effect of ATP on the hexokinase-PFK reaction (v_1) using K_i and n as the inhibition constant and cooperativity coefficient for ATP respectively:

$$f(ATP) = \left[1 + \left(\frac{[ATP]}{K_i}\right)^n\right]^{-1},$$

- Metabolites are distributed homogenously in the cytosolic and external medium.

2.3.1.2 Transduction of oscillations

The model can used to study the mechanism of intracellular propagation of nonlinear dynamics. It is reasonable to assume that when nonlinear dynamics are transmitted down the main backbone of the glycolytic pathway, the amplitude of substrates should be greater than that of the products it produces (Richard et al. 1996), i.e. each enzymatic step dampens the oscillations. Not surprisingly, a series of simulations has shown that oscillations in glycolysis can be transmitted throughout the cell via the cofactors ADP and NAD.

As a follow up, Wolf and coworkers (Wolf et al. 2001b) examined whether it is possible for cells to synchronize their oscillating dynamics if oscillations are not propagated through the backbone of the glycolytic pathway. To this end, cells of yeast with identical kinetic capabilities, but different concentrations of metabolites, were coupled together via a shared extracellular Acld pool. In Figure 2, we have simulated the coupled dynamics of three such cells. As it can be seen in Figure 2a, the cells oscillate at the same frequency but with different amplitudes and phases. Gradually, the phase shift disappeared, and in less than 20 minutes the dynamics of the cells were completely synchronized (Figure 2b).

2.3.2 Quorum sensing

Many bacteria synchronize the activation of particular functions by communicating their local cell density to each other through autoinducer (AI) molecules, an effect called “quorum sensing” (Fuqua1994). As the cell population increases, the AIs accumulate in

the surroundings, eventually reaching a critical concentration causing the differential expression of certain sets of genes, e.g. genes involved in bioluminescence. Experiments have shown that several necessary processes to bacterial colonization and virulence such as biofilm formation, bioluminescence, type III secretion, and secretion of virulence factors are regulated via quorum sensing (Fuqua1996, McFallNgai2000, Miller2001, Hammer2003, Henke2004, Waters2005).

The machinery of the gene regulatory networks that produce the AIs, detect the AIs, and respond to the changes in AI concentration differs depending on the bacteria. *Vibrio harveyi* and *Vibrio cholerae* use sensors at the membrane to track changes in the AI concentration (Freeman2000, Miller2001, Miller2002, Mok2003, Henke2004), while the AIs diffuse through the membrane and form a complex with a particular protein necessary for gene activation in *Vibrio fischeri* and *Pseudomonas aeruginosa* (Fuqua1994, Fuqua1996, Fuqua2002, Pesci1999, McKnight2000, Pearson1999). The latter type of quorum sensing bacteria will be the focus of this instructional example.

2.3.2.1 Quorum sensing model for *Vibrio fischeri*

We present a model of quorum sensing in *Vibrio fischeri* introduced by James et al. (James2000). *Vibrio fischeri* is a gram-negative bioluminescent marine bacterium that uses acyl-homoserine lactones as its AIs to directly control the luminescence (*lux*) operon (Eberhard1981, Engebrecht1983, Engebrecht1984, Fuqua2001). The model tracks the concentration of AIs (A), the concentration of the protein LuxR that the AI forms a complex with (R), and the concentration of the AI-protein complex (C) (see Figure 3). The first interaction, AI forming a complex with LuxR, is described by the binding rate

constant k_1, and the complex can break apart with dissociation rate constant k_2, giving the reactions:

$$A + R \xrightarrow{k_1} C \quad \text{and} \quad C \xrightarrow{k_2} A + R\,.$$

The resulting three differential equations for A, R, and C are

$$\frac{dA}{dt} = k_2 C - k_1 AR\,,\ \ \frac{dR}{dt} = k_2 C - k_1 AR\,,\ \ \frac{dC}{dt} = k_1 AR - k_2 C\,.$$

Through binding to the *lux* box, the complex (C) is responsible for promoting the production of the *lux* operons, which include the genes responsible for light production, *luxCDABEG*, the gene responsible for producing the AI, *luxI*, and the gene that translates into the protein AI complexes with *luxR*. When the concentration of the complex is (low) high, the *lux* box is predominantly (un-)occupied. This is account for by including a term

$$\textit{lux} \text{ box occupancy} = \frac{fC}{1 + fC}$$

where f is a proportionality constant.

Since the complex promotes the transcription of *luxI* and *luxR*, the rates of transcription are proportional to the *lux* box occupancy time. The model does not explicitly include the translation step of *luxI* and *luxR* into LuxI and LuxR, or the direct synthesis of the AI from LuxI. Instead, it is assumes these reactions to be integrated into an additional proportionality constant times the occupancy of the *lux* box.

$$\text{LuxR synthesis rate} = q\frac{fC}{1 + fC}, \qquad \text{AI synthesis rate} = p\frac{fC}{1 + fC}$$

The differential equations for A and R are thus updated to be:

$$\frac{dA}{dt} = k_2 C + p\frac{fC}{1 + fC} - k_1 AR$$

$$\frac{dR}{dt} = k_2 C + q\frac{fC}{1+fC} - k_1 AR$$

$$\frac{dC}{dt} = k_1 AR - k_2 C$$

The LuxR protein concentration is naturally reduced via enzymatic degradation and cellular volume changes from cell replication at a rate proportional to the LuxR concentration, and the chemically stable AIs freely diffuse through the cell membrane into the surrounding environment at a rate assumed to be proportional to the cellular AI concentration:

$$\text{Degradation rate of LuxR} = bR, \quad \text{Diffusion rate of AI} = nA$$

Including this effect, the differential equations for A, R, and C then become:

$$\frac{dA}{dt} = k_2 C + p\frac{fC}{1+fC} - k_1 AR - nA$$

$$\frac{dR}{dt} = k_2 C + q\frac{fC}{1+fC} - k_1 AR - bR$$

$$\frac{dC}{dt} = k_1 AR - k_2 C$$

Finally, an external concentration of AI (A_{ex}) generated by a colony of bacteria can be added to the model by including a forward rate of diffusion of AI proportional to the external concentration. This only modifies the equation for A by adding an nA_{ex}:

$$\frac{dA}{dt} = k_2 C + p\frac{fC}{1+fC} - k_1 AR - n(A - A_{ex})$$

2.3.2.2 Model analysis

To illustrate how the system of three coupled differential equations can exhibit quorum sensing behavior, we solve for the time series solutions of the differential

equations for two values of A_{ex} (A_{ex} = 1 m/l^3 and A_{ex} = 50 m/l^3). The low value of A_{ex} corresponds to the low cell-density limit where the external concentration of AIs from surrounding bacteria is minimal. Figure 4 shows the cellular concentrations of AI (dashed line) and the LuxR-AI complex (sold line). Both molecules are given initial concentrations of 1 m/l^3. The system quickly reaches steady-state conditions where the internal AI concentration matches the external one, and the concentration of the LuxR-AI complex drops to a minimal value. Since the LuxR-AI complex is responsible for activating luminescence, this situation corresponds to a dark colony. Upon increasing the concentration of external AI (corresponding to high cell density), the LuxR-AI complex is able to reach a considerably larger steady-state concentration. Since the threshold concentration of the LuxR-AI complex necessary for light production is not known, the results in Figure 5 serve as an illustration of the cell's response to a large increase in external AI concentration.

Using the model for quorum sensing in *Vibrio fischeri* proposed by James et al. (James2000), it is clear that this relatively simple set of coupled differential equations is capable of exhibiting a quorum sensing-like response when the concentration of external AI is changed. Other models exist that include more interactions in the genetic regulatory network (Muller2006, Kuttler2007). There are also models of *Pseudomonas aeruginosa*, a similar quorum sensing bacteria to *Vibrio fischeri* (Ward2001, Dockery2001).

2.4 Tools

ODEs can easily be solved by general scientific and engineering software such as matlab (www.mathworks.com) and mathematica (www.wolfram.com). Many programs have been developed primarily to facilitate the modeling of dynamical systems:

- Virtual cell (www.vcell.org)
- E Cell (www.e-cell.org)
- CellDesigner (www.celldesigner.org)
- Karyote (biodynamics.indiana.edu/CellModeling)
- MathSBML (www.sbml.org/Software/MathSBML).

There are also a number of databases, such as www.siliconcell.net, were metabolic models are stored.

3. Individual-based modeling

3.1 Background

The history of individual-based modeling, also often called agent-based modeling (ABM), goes back to the late 1940s and early 1950s work by John von Neumann where he invented cellular automata (CA). CA are most frequently simulated on finite grids, and the state of a grid-cell's neighbors is used to determine its state for the next time step. In a simple 1D example, only two states (0 or 1) are available per cell, and the CA update rules would then determine for which of the 8 possible states a cell would change its value.

Individual-based models (IbMs) were suggested in the 1980s as a possible method for studying social systems on a computer. Differently from the CA, the IbMs are typically not occupying all available grid cells and, in fact, need not be based on a grid at all.

However, similarly to the CA, each entity carries with it a pre-destined set of rules that it acts upon after polling its local environment. Due to the rapid increase in computational power for desktop PCs, IbMs started receiving serious attention in the 1990s (Wimpenny, 1997; Kreft, 2001; Pizarro, 2001).

In the following, we will enlist the IbM framework to model microbial communities, and the agents will represent individual cells, being either bacteria, archaea, or single-cell eukaryotes. Contrasting the IbM framework with that of the rate equation approach, we quickly see that the chasm in representation can be bridged. For instance, one can imagine that the internal rule-set for an agent is based on monitoring the output of a set of rate equations, such as growth, internal ATP concentration, or auto-inducer concentration in quorum sensing. However, the computational cost of including a highly detailed internal description should be measured carefully against the feasible number of simultaneous agents and the duration of the simulation.

3.2 Theory and Methodology

When a system is comprised of many agents whose interactions generate system-level dynamics that cannot be explained by their individual properties (emergent behavior), individual-based modeling is well suited for simulating the system function. Typically, IbMs of microbial communities are simulated on 2D or 3D grids where a single entity occupies a grid-cell. Before taking on the task of designing or implementing an IbM, it is necessary to clearly define the contents and scope of the project. Important questions to clarify include:

- How many species will exist in the system?

- Will the microbes be allowed to move?
- What will be the inputs and outputs of each microbe?
- How much will a microbe eat before it divides?
- After cell division, how will the (now) two cells be placed?
- Which boundary conditions will be chosen (e.g. hard walls, nutrient reservoir)?
- Which metabolic strategies, e.g. dormant or growing maximally, may be used?
- How will the microbes interact; through competition for nutrients or through more direct channels, e.g. quorum sensing, physical contact, or production of toxins?

Additionally, it is necessary to decide how nutrients and other chemicals will move in the system, as well as the shape and function of the system boundaries. In the modeling of biofilms, nutrient levels are sometimes chosen to be fixed along one of the system boundaries to simulate the presence of a reservoir, while a different boundary is chosen to be impermeable to both nutrients and cells, emulating a hard surface such as a wall.

The basis for any IbM is the set of "behavioral" rules that each microbe may follow. For every time-increment, each microbe is visited and taken through the list of possible rules. In simple cases the rule set is deterministic: whenever the local conditions are identical, a given outcome is repeated. For more sophisticated models, the microbe may choose among the available strategies with a probability that depends on past history, the local environment, or both. While implemented behavioral rules frequently have been discrete in nature, this is not a requirement of the modeling approach. For instance, a common choice in calculating the growth of a microbe from one time-point to the next is to increment an "energy storage" variable with a fixed amount. However, one may

alternatively describe the growth (rate) using Michaelis-Menten, or even double-saturation kinetics (Xavier 2007).

It is in the selection of behavioral rules that IbM intersects with game theory. In simple IbMs, the rule set only allows for interactions through the use of nutrients or occupation of space (e.g. a microbe is not allowed to grow when adjacent grid-cells are occupied). However, microbes may cooperate or compete through the production of chemical signals (quorum sensing) and toxins (West 2006). It is relatively straightforward to include a wide variety of competitive or cooperative behaviors in the behavioral rules. For instance, we can generate a class of cooperative microbes simply by lowering their possible growth rate while they produce a beneficial by-product, such as extracellular polymeric substance (EPS) or a molecule that aids the function of a different microbial species. The competing behavioral class of "cheaters" will be allowed to avoid this burden (e.g. no EPS production) and can grow at the maximal rate. In such a scenario, it is possible either for the cooperators or the cheaters to have the highest fitness, depending on growth conditions and the structure of the environment (Kreft, 2004; Xavier, 2007).

When designing an IbM, it is also necessary to carefully consider how the nutrients are distributed. In the simplest models, nutrient concentrations are chosen to be constant, while more complex realizations include discretized differential equations for the diffusive nutrient transport. These hybrid methods, combining IbM dynamics for the microbes with differential equations for the nutrients, have given highly detailed insights into the dynamics of biofilms (see e.g. Chambless 2006 for example of 3D simulation). In these approaches, it is beneficial to utilize the difference in time-scales between

diffusion (fast process) and microbial activities (slow process) such as growth. The following two-step iterative process is frequently used: (1) calculate the quasi-steady state solution for the diffusive molecules, and (2) use the identified local concentrations as input for the microbial IbM dynamics. Assuming that both microbial locations and their uptake and production rates are fixed, we may easily find the steady-state solution of the diffusion equations of, e.g. oxygen, glucose, and an auto-inducer. Note that, the microbes may act as both sinks (consumption of nutrient) and sources (production of signaling molecules).

Alternatively, we may consider the nutrients and other chemicals as discrete particles that conduct independent random walks, e.g equal probability of moving to an adjacent site, and multiple nutrient particles are allowed to occupy the same grid-site. In this representation, the effective diffusion coefficient is determined by the number of steps in the walk. Fluid flow may be incorporated by biasing the direction of the random walk. Note that one must conduct the random walk step for all particles before updating the microbial states.

3.3 Example

In a simple, deterministic 2D system where the only interaction between the microbes is competition over nutrients and available space, the rule set is:

1) Nutrient uptake.
    a) If amount of nutrient $E>e$ available in current and adjacent grid-cells, eat amount e. Add to internal energy storage: $w \rightarrow w+e$ (and appropriately subtract from E).
    b) If not, maintenance cost $m<e$ is deducted: $w \rightarrow w-m$

2) Duplication or sporulation.

   a) If at least one adjacent grid-cell is empty and internal energy storage $w>W$ (the duplication threshold) generate copy and set $w \rightarrow (w-W)/2$ in both microbes.

   b) If internal energy storage $w<T$, the sporulation threshold, microbe is inactive until nutrient level in current grid-cell $E>e$.

Naturally, we choose $T << W$. In this simple example, we are inhibiting the movement of nutrient particles, similar to microbial growth on an agar plate. By allowing for the movement of nutrients, either as a random walk of discrete particles or by differential equations (diffusion), this simple IbM can be changed to describe biofilm growth in a liquid medium. Typical initial conditions start from either a single or multiple identical microbes in the middle of the grid or along a boundary. Multiple species are simply included by e.g. changing the uptake amount from being a global constant e, to become species dependent e_s.

We can create cooperative behavior by modifying e.g. behavioral rule 1.a as follows:

1.a') If amount of nutrient $E>e$ available in current and adjacent grid-cells and majority of adjacent grid-cells occupied, eat amount $e'=e-\delta>0$. If majority of adjacent grid-cells empty, eat $e'=e$. Add to internal energy storage: $w \rightarrow w+e'$ (and appropriately subtract from E).

This straightforward rule change forces microbes to behave altruistically by taking less of the nutrients when in a dense neighborhood, and thus, improve sharing of resources.

Figure 6 shows a snapshot of a biofilm simulation of two species competing over the same food source. In addition to rules 1 and 2, we have included nutrient diffusion using the random walk approach. It is not surprising that the fast growing species (dark gray) is

dominating over the slower growing species (light gray) in the major bloom: the further away from the bottom layer (the wall) an individual is, the more nutrients are available and it can grow faster.

3.4 Tools

Several consortia have made available general purpose IbM models. The most popular open-source implementations are Swarm (www.swarm.org) and Netlogo (ccl.northwestern.edu/netlogo). A listing of available IbM software packages is available at www.swarm.org/index.php?title=Tools_for_Agent-Based_Modelling. Programs specifically tailored to microbial communities include BacSim (Kreft 2001), which is based on the Swarm toolkit, and BacLAB (Hunt, 2003).

4. Population Dynamics

4.1 Background

Population dynamics in community-level modeling comprises a coarse-grained approach compared to the two previous sections, where the focus has shifted from individual microbe to the species as basic unit. Population-level interactions between different species (macroscopic) can naturally be considered as effective per-capita rates resulting from the interacting individuals (microscopic). Thus, population interactions naturally arise from shared ecological niches and diverse metabolic capabilities of the constituent microbes.

A conventional way of classifying pair-wise population interactions is based on their effects on growth (see Table 2). The presence of one species may be beneficial [+],

detrimental [-] or neutral [0] to the other. In fact, all possible combinations of effects are observed in nature, both the symmetric (reciprocal) interactions of mutualism [++] and competition [--], and the asymmetric cases of ammensalism [0-], commensalism [0+], predator-prey or parasitism [+-]. However, this scheme does not reflect the microscopic origin of interactions. Simple abstractions of an interaction may be insufficient to quantitative analyses, and it is important to carefully consider the microscopic origin of interactions. We also note that these classification schemes constitute an idealization: In practice, the behavior of a mixed community is likely the combination of multiple interactions, often with opposing effects. A situation that is common to microbial communities consists of two (or more) species in a mixed population that compete for the same nutrient source while, at the same time, being physiologically coupled in a commensal way. Thus, we should not expect that the resulting dynamics will be predictable by "effective" uniform interaction models.

Population level descriptions provide insights that are otherwise overlooked in microscopic studies. Microbial communities from compost, the bovine rumen, acid mine drainage, and hot springs are just a few among recently studied systems that will benefit from quantitative modeling.

4.2 Theory and Methodology

4.2.1 Lotka-Volterra model and its deterministic variations

Since the early modeling of the predator-prey ecosystem, the Lotka-Volterra (LV) model (Lotka 1925, Volterra 1926) has been the *de facto* standard template for modeling mixed populations. Though LV had originally aimed at modeling the specific case of

predator-prey system, its current usage has been expanded past the predator-prey setting to include positive interactions. In its simplest version, the population size of a prey (x_1) and its predator (x_2) satisfy the following set of nonlinear differential equations.

$$\frac{\mathrm{d}}{\mathrm{d}t}x_1(t) = \alpha x_1 - \beta x_1 x_2 \quad , \qquad \frac{\mathrm{d}}{\mathrm{d}t}x_2(t) = \gamma x_1 x_2 - \delta x_2$$

Here α and δ are the growth and decay rates for the prey and predator populations, unaffected by the negative (predation) inter-species interaction. The coefficients β and γ represents the strength of the detrimental and beneficial effects on prey and predator population owing to the predation. Due to the particular functional form of these equations, the Jacobian of this system has purely imaginary eigenvalues, regardless of the parameter combinations. Consequently, the two-species LV system has sustained oscillatory behavior with a characteristic frequency of $\sqrt{\alpha\delta}\,/\,2\pi$.

The exponential growth of prey population has been a target for modifications. The original LV assumes no resource limits, which oftentimes is unrealistic. To include the resource-mediated intra-species competition, we require a negative term that would counterbalance exponential growth. Thus introduced is the logistic growth rate, $\alpha n(1 - n/K)$, where r is the per-capita growth rate and K is the carrying capacity of the ecosystem for species i. The modified LV with the logistic growth with finite carrying capacity for the prey population is now

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}n_1(t) = \alpha n_1\left(1 - \dfrac{n_1}{K}\right) - \beta n_1 n_2 \\ \dfrac{\mathrm{d}}{\mathrm{d}t}n_2(t) = \gamma n_1 n_2 - \delta n_2 \end{cases}$$

which has the two nontrivial (excluding $x_1 = x_2 = 0$) steady states (Figure 7):

$$(x_1, x_2) = (K, 0),\ \left(\frac{\delta}{\gamma}, \frac{r}{\beta}\left[1 - \frac{\delta}{\gamma K}\right]\right).$$

The first solution corresponds to predator extinction and prey proliferation, which is stable as long as $K < K_c \equiv \delta/\gamma$ (the extinction threshold). Stable population coexistence (second solution) is possible only when $K > K_c$. Linear stability analysis further shows that coexistence is either a stable node or a focus, and no oscillatory behavior is expected unless the carrying capacity diverges (Mobilia et al. 2007).

We may generalize the LV population model (which we will refer to as GLV) to include competitive interactions among species by adding an extra, negative term following the spirit of mass-action:

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i(t) = x_i\left(r_i - \sum_{j=1}^{n} A_{ij}x_j\right),$$

where n is the total number of interacting species. The diagonal elements $A_{ii} > 0$ can be identified (to a multiplicative constant) with the inverse of the carrying capacity of species i. The off-diagonal elements $A_{ij} > 0$ represent the strength of j's negative effect on i, which is related to the distance between the two species in niche space.

Finally, a unified scheme for the community interactions is obtained by removing the positivity constraint on the off-diagonal elements A_{ij} in GLV. The majority of studies on mutualistic interactions have been using this representation as a template framework. However, all eigenvalues of the interaction matrix must have positive real parts for the system to be stable. High-diversity communities tend to become unstable as the interaction network becomes more complex, reminiscent of the work by May in the 1970s (May 1973, 1976). Recent studies have revisited this problem and found potential

positive effects of complexity: High-diversity, stable LV systems arise if the interaction network evolves flexibility through adaptive behavior (Kondoh 2003, Ackland & Gallager 2004).

4.2.2 Effects of spatial heterogeneity

In general, microbial populations are spatially heterogeneous and not well-stirred "bioreactors" as assumed in the original LV work. Even marine microbes aggregate in the search for food using chemotaxis. We may introduce spatial structure into the deterministic framework by using an embedding space, where the individuals move around in the search for food and shelter. Interaction effects are no longer instantaneous, leading to time delays that stabilizes the community (Murray 2002, Collet & Eckmann 1990). A natural extension of LV to allow for the random movement of cells is by way of diffusion terms, turning the LV into the coupled partial differential equations

$$\begin{cases} \dfrac{\partial n_1(\mathbf{x},t)}{\partial t} = D_1\nabla^2 n_1(\mathbf{x},t) + \alpha n_1(\mathbf{x},t) - \beta n_1(\mathbf{x},t)n_2(\mathbf{x},t) \\ \dfrac{\partial n_2(\mathbf{x},t)}{\partial t} = D_2\nabla^2 n_2(\mathbf{x},t) + \gamma n_1(\mathbf{x},t)n_2(\mathbf{x},t) - \delta n_2(\mathbf{x},t) \end{cases}$$

where D_i is the diffusion coefficient of species i. For the case of two-species competition, this coupled reaction-diffusion system is known to contain propagating wave-front solutions in 1D of the form $x_i(t) = f(x_i - v_i t)$ that interpolate between the two steady states identified above. Convergence to the steady state monotonically or with oscillations depends on the choice of rate parameters (Murray 2002).

4.2.3 Stochastic modeling

Randomness is a defining character of population processes, often diverting the dynamics from deterministic predictions. Depending on the origin of the "noise," population stochasticity may be classified by the following categories:

- Within-individual variability ("demographic stochasticity"),
- Cell-to-cell variability and age structure,
- Spatial heterogeneity (quenched or annealed),
- Temporal fluctuation of environment.

The first two categories stem from the random timing of birth-death events and the discrete nature of individuals. These factors play a lesser role as the population grows in size, but may still have significant local effects. In fact, local extinctions commonly occur in nature, which is consistent with observations in stochastic simulations. The latter two categories are extrinsic in origin and can be described in terms of quenched or annealed noise. Note that, contrary to intrinsic noise, there is no constraint on the noise amplitude or temporal correlations. Overall, the different sources of noise work together in real ecosystems, and interesting behaviors emerges from their combinatorial effects (Kussel & Leibler 2005, Thattai & van Oudenaarden 2004). Given a non-interacting single population with a discrete phenotypic distribution, the time-evolution of species n_i can be described by the following matrix equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}n_i(t) = r_i\big(E(t)\big)n_i(t) + \sum_j T_{ij}\big(E(t)\big)n_j(t)$$

where $E(t)$ is the random discrete variable representing the environment at time t. $r_i(E(t))$ is the environment-dependent fitness of phenotype i, and the matrix elements $T_{ij}(E(t))$ are transition probabilities for an individual to switch from the j-th tothe i-th phenotype.

Interestingly, maximal growth occurs when the phenotypic switching rate is similar to that of environmental fluctuations, and random switching outperforms responsive switching if environmental changes are slow or mostly predictable.

4.3 Example

4.3.1 Marine phage community

Recent work on marine phage communities demonstrate how the general framework of LV can be improved, and the importance of investigating microscopic origins of population growth. Hoffmann and colleagues (Hoffmann *et. al*, 2007) studied the interaction of marine phages (predator) and their host microbes (prey) by modeling the multispecies community as a simple predator-prey model. This can be justified since the phage-host interaction is highly specific and the dominant population effectively is representative of the overall community (Thingstad 2005).

The key observation from this approach is that the observed cooperativity is casued by spatio-temporally non-uniform nutrient condition ascribed to a colloid-type organic detritus called "marine snow." The marine snow enhances aggregation of microbes and their predators, generating a positive feedback loop: The clustering around discrete food sources leads to locally high concentrations of lysed host cells, that further attract more predators. The consequence is a superlinear dependence of predation rate in the phage population, represented as a quadratic interaction in the phage density:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} n_1(t) = \alpha n_1 - \beta n_1 n_2^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} n_2(t) = \gamma n_1 n_2^2 - \delta n_2^2 \end{cases}$$

The population dynamics predicted by this model follow experimental data closely.

4.3.2 Identification of unknown species interactions

Intra- and inter-species interactions among microbes are mainly responsible for the ripening process in spreadable cheeses. A recent study used the population dynamics approach to identify interactions in a spreadable cheese bacteria-eukaryote community composed of six bacteria and three yeast species (Mounier *et. al*., 2008). The bacterial population behavior could be grouped into two quasi-species, resulting in a five-species model system. Using the GLV formulation as a starting point, entries in the interaction matrix A were selected to give simulated population dynamics that agreed with measurements. The identified possible realizations of A were further narrowed down through a species-removal study: a single quasi-species was removed at a time, and population dynamics for the remaining species were measured. As a result, the web of interaction between the five groups could be identified (see Figure 8). Considering the experimental difficulties in resolving interspecies interactions in strongly interacting communities, the GLV modeling approach provides a useful first step.

4.4 Tools

4.4.1 General-purpose ODE solver

The SBML ODE solver library (SOSlib) is a programming library for formula representation to construct ordinary differential equation (ODE) systems, their Jacobian matrix, a parameter dependency matrix and other derivatives in the Systems Biology Markup Language (SBML). SOSlib provides efficient interfaces to well-established methods in theoretical chemistry, biology and systems theory

http://www.tbi.univie.ac.at/~raim/odeSolver

4.4.2 Stochastic simulator

Dizzy is a software for stochastic chemical relations simulation. It provides a model definition, implementation of several stochastic and deterministic algorithms, a graphical display of a model. It is a standard free software written in Java and is supported on Windows XP, Fedora Core 1 Linux, and Macintosh.

http://magnet.systemsbiology.net/software/Dizzy

Acknowledgments

This work performed under the auspices of the DoE by LLNL under Contract DE-AC52-07NA27344, and supported by the LLNL Laboratory Directed Research and Development program on grant 06-ERD-061.

Tables

Model differential equations
$\frac{d[Glc]}{dt} = I - k_1[Glc][ATP]f(ATP)$
$\frac{d[FBP]}{dt} = k_1[Glc][ATP]f(ATP) - k_2[FBP]$
$\frac{d[TRP]}{dt} = 2k_2[FBP] - v_{4\&5} - k_3[TRP][NADH]$
$\frac{d[3PG]}{dt} = v_{4\&5} - k_7[3PG]$
$\frac{d[Pyr]}{dt} = k_7[3PG] - k_8[Pyr]$
$\frac{d[Acld(c)]}{dt} = k_8[Pyr] - k_9[NADH][Acld(c)] - X$
$\frac{d[Acld(m)]}{dt} = \frac{V}{V_m^*}X - k_{10}[Acld(m)]$
$\frac{d[ATP]}{dt} = v_{4\&5} - 2k_1[Glc][ATP]f(ATP) + k_7[3PG] - k_6[ATP]$
$\frac{d[NADH]}{dt} = v_{4\&5} - k_9[NADH][Acld(c)] - k_3[TRP][NADH]$

*V_m=*volume of the medium*

Table 1. List of differential equations for a simplified model of glycolysis (Wolf 2000)

Interaction mode	Reciprocity	Cell-cell direct contact	Sign of interactions
Mutualism	Y	N	[++]
Competition	Y	N	[--]
Commensalism	N	N	[+0]
Ammensalism	N	N	[-0]
Predator-prey, parasitism	N	Y	[+-]

Table 2. Overview of species-level interaction classes in population based modeling.

Figure captions

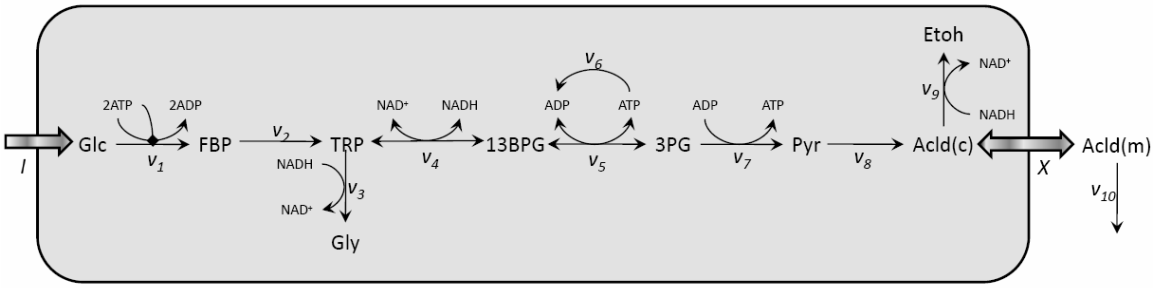

Figure 1. Schematic diagram of anaerobic glycolysis. Glc=glucose, TRP=triose-phosphates. v_1=hexokinase & PFK, v_2=aldolase, v_3=glycerol-3-phosphate dehydrogenase & glycerol kinase, v_4=GAPDH, v_5=PGK, v_6=ATPase, v_7=phosphoglycerate mutase & enolase & pyruvate kinase, v_8=pyruvate decarboxylase, v_9=alcohol dehydrogenase, v_{10}=degradation of Acld.

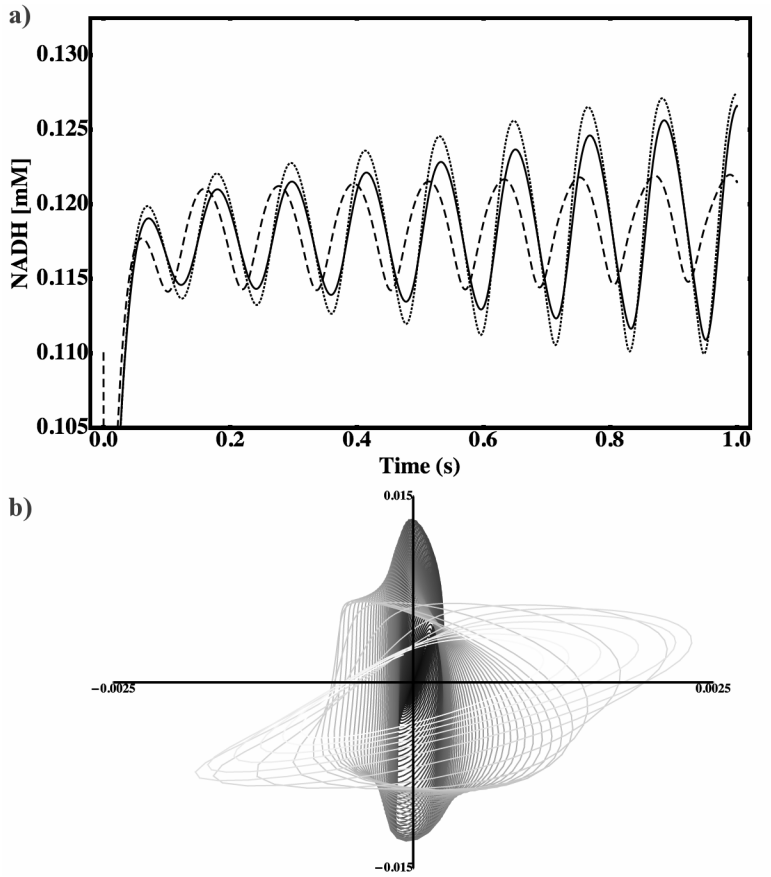

Figure 2. Coupled dynamics of glycolysis in 3 cells with identical kinetic capabilities but different starting metabolite concentrations. **a)** Oscillating concentration of NADH. The oscillation frequency is the same for all three cells while the starting amplitudes (Am) and phases differ ($Am_A > Am_B > Am_C$). **b)** Amplitude differences in NADH oscillations between two pairs of cells (A-B and B-C). Time course is represented by shading (early=white, late=black).

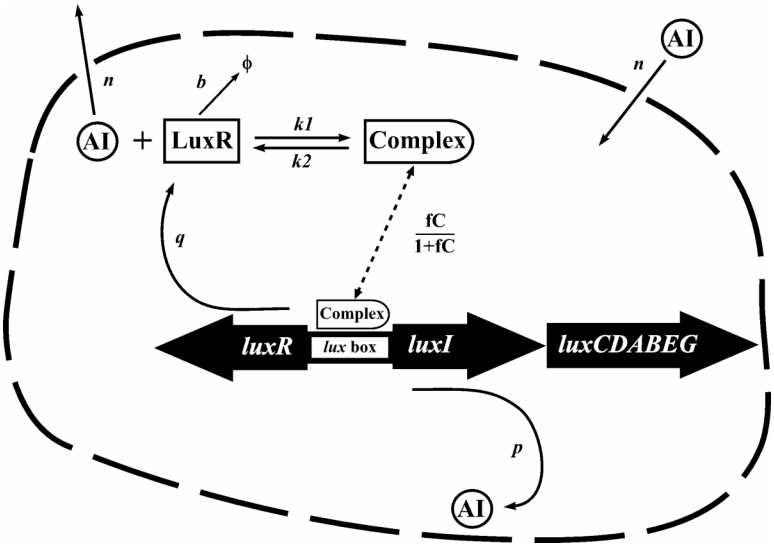

Figure 3. Schematic of quorum sensing network used by *Vibrio fischeri* to regulate luminescence. AI (A) binds the protein LuxR (R) to form complex (C) with a forward rate of k_1 and a dissociation rate of k_2, and diffusion of AI through cellular membrane with constant *n*. LuxR is degraded at a rate *b*. The C complex occupies the *lux* box proportional to *fC / (1 + fC)* and promotes the transcription of *luxR*, *luxI*, and *luxCDABEG* with rate *q*. AI is produced at rate p from LuxI.

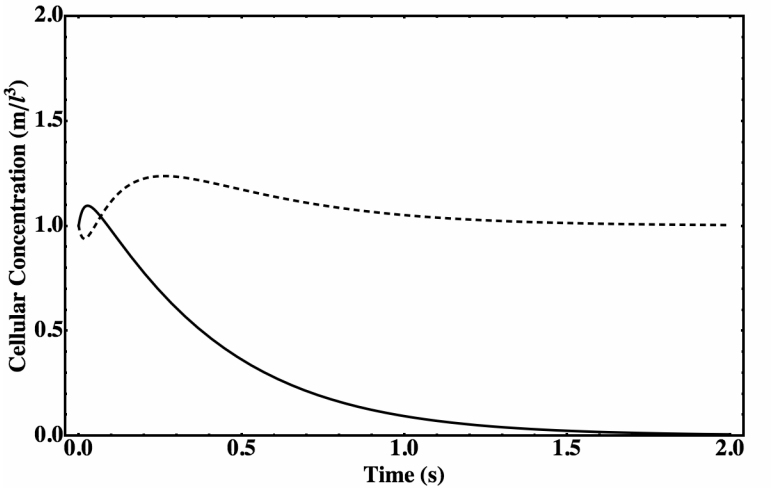

Figure 4. Low cell density response of LuxR-AI complex (solid line) and AI (dotted line) concentrations. Starting concentrations are 1 m/l^3 for LuxR-AI, AI, and AI_{ext}. The system quickly reaches its steady-state values where LuxR-AI complex concentration is minimal and the AI concentration matches AI_{ext}. The other parameters are: $k_1 = 25$ (l^3 m^{-1} t^{-1}), $k_2 = 10$ (t^{-1}), $n = 10$ (t^{-1}), $b = 10$ (t^{-1}), $p = 5$ (m l^{-3} t^{-1}), $q = 2.5$ (m l^{-3} t^{-1}), and $f = 0.25$ (l^3 m^{-1}).

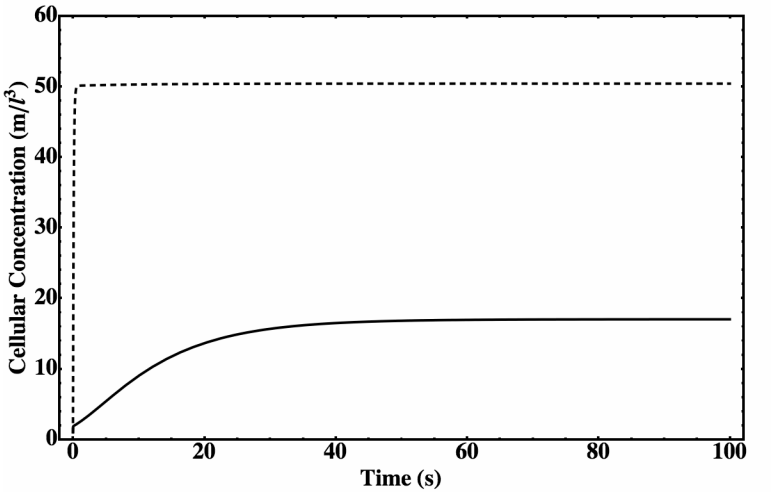

Figure 5. High cell density response of LuxR-AI complex (solid line) and AI (dotted line). Starting concentrations are 1 m/l^3 for LuxR-AI, AI, and 50 m/l^3 for AI_{ext}. The system quickly reaches its steady-state values where LuxR-AI complex concentration can initiate light production and the AI concentration matches AI_{ext}. The other parameters are as in Figure 4.

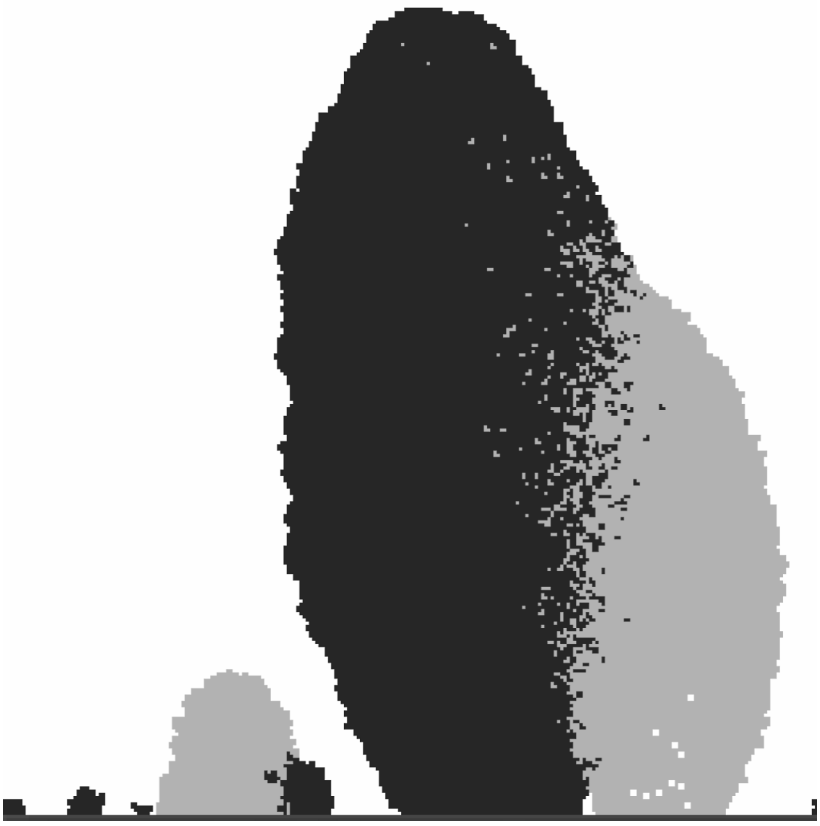

Figure 6. Simulated biofilm of two competing species growing on an impermeable boundary. Fast growing (dark gray) dominates over slower growing (light gray). Substrate gradients are generated by random walks of discrete nutrient packets.

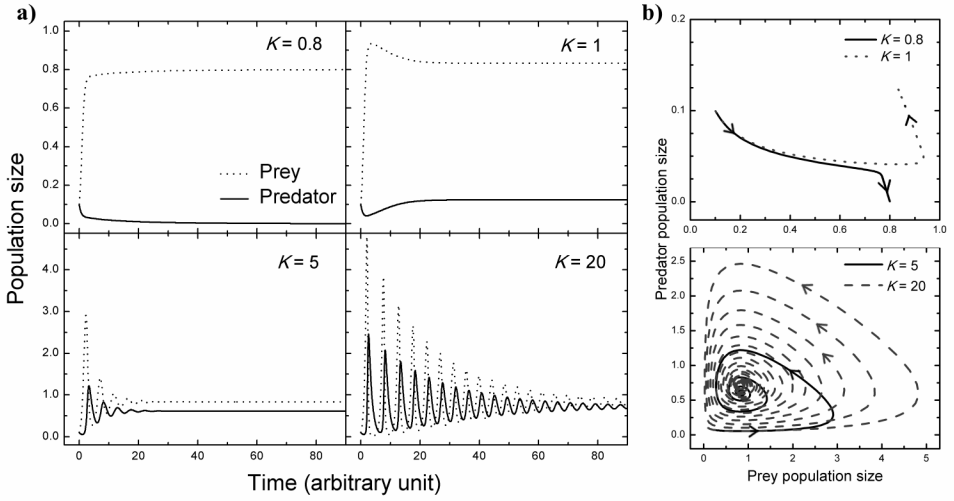

Figure 7. Competitive Lotka-Volterra (LV) dynamics. (a) Time evolution of the population size from Eq. (2). All the systems start $n_1(0) = n_2(0) = 0.1$ (arbitrary units) and the time scale is set in units of $1/\delta$ (~predator's lifespan). Rate parameters α=2.3, β=3.1, γ=1.2, and the carrying capacity K is varied from 0.8 to 20 (K_c = 0.833). The mixed population state is stable for $K > K_c$. (b) Trajectories in n_1-n_2 space shows the attractor for different carrying capacities.

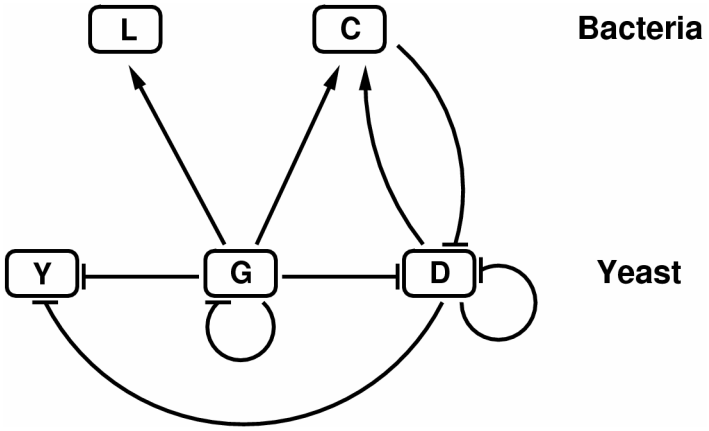

Figure 8. Interaction network identified by generalized LV (GLV) analysis. Arrows and blunt ends stand for positive and negative interactions, respectively. D, *Debaryomyces hansenii*; Y, *Yarrowia lipolytica*; G, *Geotrichumcandidum*; L, *Leucobacter sp*.; C, group containing *Arthrobacter arilaitensis*, *Hafnia alvei*, *Corynebacterium casei*, *Brevibacterium aurantiacum*, and *Staphylococcus xylosus*.